REPORT DOCUMENTATION PAGE

Form Approved
OMB No 0704-0188

1a. REPORT SECURITY CLASSIFICATION		1b. RESTRICTIVE MARKINGS			
Unclassified					
2a. SECURITY CLASSIFICATION AUTHORITY		3. DISTRIBUTION / AVAILABILITY OF REPORT			
N/A		Approved for Public Release;			
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE		Distribution Unlimited			
N/A					
4. PERFORMING ORGANIZATION REPORT NUMBER(S)		5. MONITORING ORGANIZATION REPORT NUMBER(S)			
Technical Report No. 293		AFOSR-TR· 90 0852			

6a. NAME OF PERFORMING ORGANIZATION	6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION
University of North Carolina Center for Stochastic Processes		AFOSR/NM

. ADDRESS (City, State, and ZIP Code)	7b. ADDRESS (City, State, and ZIP Code)
Statistics Department CB #3260, Phillips Hall Chapel Hill, NC 27599-3260	Bldg. 410 Bolling Air Force Base, DC 20332-6448

. NAME OF FUNDING / SPONSORING ORGANIZATION	8b OFFICE SYMBOL (If applicable)	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
AFOSR	NM	F49620 85C 0144

ADDRESS (City, State, and ZIP Code)	10. SOURCE OF FUNDING NUMBERS			
Bldg. 410 Bolling AFB, DC 20332-6448	PROGRAM ELEMENT NO	PROJECT NO.	TASK NO	WORK UNIT ACCESSION NO
	6.1102F	2304	A5	

11. TITLE (Include Security Classification)

Sampling designs for estimating integrals of stochastic processes using quadratic mean derivatives

12. PERSONAL AUTHOR(S)

Benhenni, K. and Cambanis, S.

13a. TYPE OF REPORT	13b. TIME COVERED		14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
preprint	FROM ____ TO ____		19 90, April	29

16. SUPPLEMENTARY NOTATION

None.

17.	COSATI CODES		18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	N/A
XXXXXX	XXXXXXXXXX	XXX	

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

(see over)

DTIC
ELECTE
AUG 3 0 1990
S D
D

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT			21. ABSTRACT SECURITY CLASSIFICATION
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS			UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL		22b. TELEPHONE (Include Area Code)	22c. OFFICE SYMBOL
~~Professor Eytan Barouch~~ Dr. J. H. Sjogren		(202) 767-5026	AFOSR/NM

DD Form 1473, JUN 86   Previous editions are obsolete.   SECURITY CLASSIFICATION OF THIS PAGE

The problem of estimating the integral of a stochastic process by linear estimators based on observations of the process and its existing quadratic mean (q.m.) derivatives at a finite number of sampling points is considered. The process is assumed to have exactly K q.m. derivatives, K=0,1,2,... . The asymptotic performance of optimal-coefficient estimators that depend on an inverse covariance matrix is determined for regular sampling designs under slightly different assumptions than those in Sacks and Ylvisaker (1970). Simple-coefficient estimators based on a trapezoidal rule with a correction term that depends on the q.m. derivatives of the process at all sampling points of a regular design are introduced. Their asymptotic performance is identical to that of the optimal-coefficient estimators.

CENTER FOR STOCHASTIC PROCESSES

Department of Statistics
University of North Carolina
Chapel Hill, North Carolina

SAMPLING DESIGNS FOR ESTIMATING INTEGRALS OF

STOCHASTIC PROCESSES USING QUADRATIC MEAN DERIVATIVES

by

Karim Benhenni

and

Stamatis Cambanis

90 08 27 020

SAMPLING DESIGNS FOR ESTIMATING INTEGRALS OF
STOCHASTIC PROCESSES USING QUADRATIC MEAN DERIVATIVES

by

Karim Benhenni[*]
and
Stamatis Cambanis

Statistics Department
University of North Carolina
Chapel Hill, NC 27599-3260

Abstract.

The problem of estimating the integral of a stochastic process by linear estimators based on observations of the process and its existing quadratic mean (q.m.) derivatives at a finite number of sampling points is considered. The process is assumed to have exactly K q.m. derivatives, K=0,1,2,... . The asymptotic performance of optimal-coefficient estimators that depend on an inverse covariance matrix is determined for regular sampling designs under slightly different assumptions than those in Sacks and Ylvisaker (1970). Simple-coefficient estimators based on a trapezoidal rule with a correction term that depends on the q.m. derivatives of the process at all sampling points of a regular design are introduced. Their asymptotic performance is identical to that of the optimal-coefficient estimators.

[*]Now at Department of Mathematical Sciences , Northern Illinois University, DeKalb, IL 60115

1. INTRODUCTION

We consider the estimation of the weighted integral of a stochastic process X over a finite interval

$$I = \int_a^b \phi(t)X(t)dt$$

from observations of the process at a finite number of sampling points. The process $X = \{ X(t), a \leq t \leq b \}$ has mean 0, continuous covariance function $R(s,t) = E[X(s)X(t)]$, and exactly K quadratic mean (q.m.) derivatives $(K=0,1,2,\ldots)$. The weight ϕ is a known (nonrandom) continuous function. The performance of an estimator is measured through the mean square error.

In BC (1989), we concentrated on linear estimators of the form $I_n = \Sigma_{i=0}^n c_{i,n} X(t_{i,n})$ that are based only on the observations of the process X at (n+1) sample points of a regular sequence of sampling designs $\{T_n(h) = \{t_{i,n}\}_{i=0}^n, \ a=t_{0,n}<t_{1,n}<\ldots<t_{n,n}=b\}_{n=1}^\infty$ generated by a positive continuous density h via

$$\int_a^{t_{i,n}} h(t)dt = i/n, \quad i=0,1,\ldots,n.$$

Here we consider linear estimators of I based on observations of the process together with its existing q.m. derivatives $X^{(1)}(t),\ldots,X^{(K)}(t)$ at (n+1) sampling points $T_n=\{t_{i,n}\}_{i=0}^n$ over the finite interval [a,b], and using coefficients $C_n = \{c_{i,j,n}\}_{i=0,\ldots,n}^{j=0,\ldots,K}$. They are of the form

$$_dI_n = \sum_{i=0}^n \sum_{j=0}^K c_{i,j,n} \, X^{(j)}(t_{i,n}).$$

We want to determine the asymptotic performance of these types of estimators based on regular sampling designs $\{T_n(h)\}_n$ and appropriate coefficients $\{C_n\}_n$. We are also interested in finding asymptotically optimal designs $\{T_n^*\}_n$ and

estimator-coefficients $\{C_n^*\}_n$, in the sense that

$$E(I-{}_dI_n^*)^2 \Big/ \inf_{T_n,C_n} E(I-{}_dI_n)^2 \xrightarrow[n]{} 1$$

where the infinum is taken over all sampling designs T_n with $(n+1)$ sample points and all choices of coefficients C_n. (Optimal designs for fixed sample size, when they exist, are in general hard to determine.)

When no q.m. derivatives are used, the optimal-coefficient estimators are of the form $\hat{I}_n = f'_{T_n} R_{T_n}^{-1} X_{T_n}$, where $X'_{T_n} = (X(t_{0,n}),\dots,X(t_{n,n}))$, $R_{T_n} = \{R(t_{i,n},t_{j,n})\}_{i,j=0}^n$ assumed nonsingular, $f'_{T_n} = (f(t_{0,n}),\dots,f(t_{n,n}))$ and $f(t) = \int_a^b R(s,t)\phi(s)ds$. In BC (1989) the asymptotic performance of these optimal-coefficient estimators using regular sequences of sampling designs, $\hat{I}_n(h)$, is obtained for a general class of processes with exactly K q.m. derivatives, K=0,1,2,..., namely

$$(1.1) \qquad n^{2K+2} E[I-\hat{I}_n(h)]^2 \xrightarrow[n]{} \frac{|B_{2K+2}|}{(2K+2)!} \int_a^b \frac{\alpha_K(t)\phi^2(t)}{h^{2K+2}(t)} dt,$$

where B_m is the mth Bernoulli number. In particular when the sampling density h^* is proportional to $(\alpha_K\phi^2)^{1/(2K+3)}$, then the asymptotic performance becomes

$$(1.2) \quad n^{2K+2} E[I-\hat{I}_n(h^*)]^2 \xrightarrow[n]{} \frac{|B_{2K+2}|}{(2K+2)!} \{\int_a^b [\alpha_K(t)\phi^2(t)]^{1/(2K+3)}dt\}^{2K+3} \overset{\Delta}{=} C^*$$

with minimal value C^* of the asymptotic constant. This result, together with the asymptotic optimality of $\hat{I}_n(h^*)$ was shown for K = 0 and 1 by Sacks and Ylvisaker (1966,1968,1970) and for general K by Eubank, Smith and Smith (1982) but for a more restrictive class of covariances.

Optimal-coefficient estimators which use the values of the process and of its existing q.m. derivatives at the sampling points, were considered by Sacks and Ylvisaker (1970). Their mean square errors have the same rate of convergence to zero, $n^{-(2K+2)}$, as in (1.1) and (1.2) for the

optimal-coefficient estimators which do not use the existing q.m. derivatives, but they have smaller asymptotic constant. In Section 3 we give a simple proof of this result under slightly different assumptions than those considered by Sacks and Ylvisaker (1970a) where the jump function $\alpha_K(\cdot)$ was assumed to be constant, and under additional assumptions on the weight function ϕ. The proof given here involves only a simple numerical approximation of integrals of deterministic functions using derivatives (Proposition 1).

These optimal-coefficient estimators require the inversion of a $(K+1)(n+1) \times (K+1)(n+1)$ covariance matrix, and hence they are liable to numerically instabilities for large sample size. More significantly, they require knowledge of the covariance matrix (and its partial derivatives) and thus they are not generally robust. Stein (1988) showed that if an incorrect covariance is used which is compatible (in an appropriate but restrictive sense) with the true covariance, then the asymptotic performance of the optimal-coefficient estimators under these two covariances is identical. However this condition fails to be satisfied when for instance asymptotically consistent estimators of the unknown covariance function are used. For example, the covariance function $R(t) = e^{-\theta|t|}$ of a zero mean Gaussian process and the covariance function $\hat{R}(t) = e^{-\hat{\theta}|t|}$, where $\hat{\theta}$ is an asymptotically consistent estimator of θ, are not compatible.

In order to address the issues of robustness and of knowledge of the covariance function, a sequence of estimators was introduced in BC (1989) which uses regular sampling designs $T_n(h)$ and simple coefficients (not depending on the covariance), and is based on the weighted Gregory formula. These estimators use only observations of the process (but not its existing derivatives) at the sample points. Their asymptotic performance was shown to be identical to that in (1.1) and (1.2) for the optimal-coefficient estimators for general processes with exactly K q.m. derivatives, K=0,1,2,... . These

highly nonparametic estimators are given by

$$I_n(h) = \frac{1}{n} \{ \frac{1}{2} \frac{\phi(a)}{h(a)} X(a) + \sum_{i=1}^{n-1} \frac{\phi(t_{i,n})}{h(t_{i,n})} X(t_{i,n}) + \frac{1}{2} \frac{\phi(b)}{h(b)} X(b) \}$$

(1.3)

$$- \frac{1}{n} \sum_{j=1}^{K} w_j \{ \Delta^j [\frac{\phi(t_{n-j,n})}{h(t_{n-j,n})} X(t_{n-j,n})] + (-1)^j \Delta^j [\frac{\phi(a)}{h(a)} X(a)] \}$$

where Δ^j denotes j^{th} order difference, $\Delta^j g(t_{i,n}) = \Sigma_{r=0}^{j} (-1)^{j-r} \binom{j}{r} g(t_{i+r,n})$, $0 \leq i+j \leq n$, and

$$w_j = [(-1)^j/(j+1)!] \int_0^1 t(t-1)\ldots(t-j)dt$$

for $j \geq 1$, and can be written in the form

(1.4')
$$I_n(h) = \frac{1}{n} \sum_{i=0}^{n} a_i \, (\frac{\phi X}{h})(t_{i,n})$$

where the coefficients a_i, $i=0,\ldots,n$, are given by

(1.4")
$$a_i = \begin{cases} \frac{1}{2} - \Sigma_{j=1}^{K} w_j & \text{for } i = 0, n, \\ 1 + (-1)^{i+1} \Sigma_{j=1}^{K} \binom{j}{i} w_j & \text{for } 1 \leq i \leq K, \\ 1 & \text{for } K+1 \leq i \leq n-K-1, \\ a_{n-i} & \text{for } n-K \leq i \leq n-1. \end{cases}$$

For example, the values of a_i, $i=0,\ldots,n$, for $K=0,1,2,3,4$, (and appropriately large n) are as follows:

K = 0: $\frac{1}{2}$, 1, 1, 1, 1, 1,...,1, 1, 1, 1, 1, $\frac{1}{2}$

K = 1: $\frac{5}{12}$, $\frac{13}{12}$, 1, 1, 1, 1,...,1, 1, 1, 1, $\frac{13}{12}$, $\frac{5}{12}$

K = 2: $\frac{3}{8}$, $\frac{7}{6}$, $\frac{23}{24}$, 1, 1, 1,...,1, 1, 1, $\frac{23}{24}$, $\frac{7}{6}$, $\frac{3}{8}$

K = 3: $\frac{251}{720}$, $\frac{299}{240}$, $\frac{211}{240}$, $\frac{739}{720}$, 1, 1,...,1, 1, $\frac{739}{720}$, $\frac{211}{240}$, $\frac{299}{240}$, $\frac{251}{720}$

K = 4: $\frac{95}{288}$, $\frac{317}{240}$, $\frac{23}{30}$, $\frac{793}{720}$, $\frac{157}{160}$, 1,...,1, $\frac{157}{160}$, $\frac{793}{720}$, $\frac{23}{30}$, $\frac{317}{240}$, $\frac{95}{288}$

In Section 4, we introduce simple-coefficient estimators which use the values of the process X and of its existing q.m. derivatives at all sampling points of a regular sampling design $T_n(h)$. They are based on the trapezoidal rule for integral approximation with a correction term that depends on the values of the K q.m. derivatives at all sample points, namely on the weighted Rodriguez formula (Section 2). Their asymptotic performance is shown in Theorem 2 to be identical with that of the optimal-coefficient estimators which use the values of the q.m. derivatives of X. The simple-coefficient estimators do not require precise knowledge of the covariance function R (or of its derivatives), and thus they are fairly robust. They are also numerically stable in view of their simple form. However, they are impractical for applications where q.m. derivatives of X at the sample points cannot be observed. In such cases the q.m. derivatives can be approximated by finite differences. The resulting estimators (4.2) use only the values of the process at the sampling points, and therefore cannot have better asymptotic performance than that in (1.1) and (1.2) for the simple-coefficient estimators considered in BC (1989). However, they may have comparable performance for small or moderate sample size, as the example in Section 5 illustrates.

In Section 5, we compare the finite sample size performance of the two types of simple-coefficient estimators based only on the values of a stationary process with K=2 q.m. derivatives under both asymptotically optimal sampling designs $\{T_n(h^*)\}_n$ and uniform sampling. We find that in this example the two estimators have the same asymptotic performance. We also find that the estimators (optimal and simple-coefficient) that use the existing q.m. derivatives clearly perform better than the estimators that do not for moderate and large samples.

In Section 2 we develop a quadrature formula for integral approximation of deterministic functions based on the derivatives of the integrand under regular

sampling designs. These results are used in Sections 3 and 4.

2. APPROXIMATION OF INTEGRALS OF NONRANDOM FUNCTIONS USING DERIVATIVES AND REGULAR SAMPLES.

The classical rules for approximating integrals of deterministic functions use periodic sampling designs, and some of these have been extended to regular sampling designs in BC (1989). Here we consider the trapezoidal rule with a correction term that depends on weighted derivatives of the integrand at all sampling points of the regular design. These rules are derived from the standard Rodriguez formula for periodic sampling in Shoenberg (1969), and are generalized to regular sampling in Proposition 1. Their asymptotic properties are given in Proposition 2.

We will use the "h-weighted" derivatives of a function f defined by

$$f_{(0)} = f/h, \qquad f_{(j)} = f^{(1)}_{(j-1)}/h \qquad \text{for } j \geq 1,$$

provided f and h have the required derivatives, and the h-weighted differential operator d_j:

$$d_j f = f_{(j)} \qquad \text{for } j \geq 0,$$

so that $d_j f_{(i)} = f_{(i+j)}$. Also the probability function $H(x,y)$ is defined by

$$H(x,y) = \begin{cases} \int_x^y h(t)dt & \text{for } a \leq x \leq y \leq b, \\ -H(y,x) & \text{for } a \leq y \leq x \leq b, \end{cases}$$

and satisfies for all $x \neq y$ and $j \geq 0$,

$$\frac{\partial}{\partial x} H(x,y) = -h(x), \qquad \int_x^y H^j(t,y)h(t)dt = \frac{1}{j+1} H^{j+1}(x,y).$$

In the particular case where h is the uniform density on [a,b]: $h(t) = (b-a)^{-1}$, then $f_{(j)} = (b-a)^{j+1} f^{(j)}$, $j \geq 0$, and $H(x,y) = (y-x)/(b-a)$.

<u>Proposition 1</u>. (<u>Weighted Rodriguez formula for regular sampling</u>). If f and h

have $m(\geq 1)$ continuous derivatives on $[a,b]$, and if $\{T_n(h)\}_{n=1}^{\infty}$ is a regular sequence of sampling designs, then we have

$$\int_a^b f(t)dt = \frac{m!}{(2m)!} \sum_{j=0}^{m-1} \frac{1}{n^{j+1}} \frac{(2m-j-1)!}{(m-j-1)!(j+1)!} \sum_{i=0}^{n-1} [f_{(j)}(t_{i,n}) + (-1)^j f_{(j)}(t_{i+1,n})]$$

$$+ \frac{1}{(2m)!} \sum_{i=0}^{n-1} \int_{t_{i,n}}^{t_{i+1,n}} d_m[H(t_{i,n},t) \, H(t,t_{i+1,n})]^m \, f_{(m)}(t)h(t) \, dt.$$

Furthermore, if f and h have $2m$ continuous derivatives on $[a,b]$, then the remainder integral term can be written as

$$\frac{(-1)^m}{(2m)!} \sum_{i=0}^{n-1} \int_{t_{i,n}}^{t_{i+1,n}} [H(t_{i,n},t)H(t,t_{i+1,n})]^m \, f_{(2m)}(t)h(t) \, dt.$$

<u>Proof</u>. Fix $[x,y] \subset [a,b]$, and define the weighted Legendre polynomial

$$P_{(0)}(t) = \frac{1}{m!} \, d_m[H(x,t) \, H(t,y)]^m .$$

It can also be expressed as follows

$$P_{(0)}(t) = \frac{1}{m!} \, d_m[H(x,t)(H(x,y) - H(x,t))]^m$$

$$= \frac{1}{m!} \sum_{i=0}^{m} \binom{m}{i}(-1)^i \, H(x,y)^{m-i} \, d_m[H(x,t)]^{m+i}$$

$$= \sum_{i=0}^{m} \frac{(m+i)!}{(m-i)! \, i!^2} (-1)^i \, H(x,y)^{m-i} \, H(x,t)^i ,$$

from which we obtain for $0 \leq j \leq m$,

$$P_{(j)}(t) = \sum_{i=j}^{m} \frac{(m+i)!}{(m-i)! \, i!} \frac{(-1)^i}{(i-j)!} \, H(x,y)^{m-i} \, H(x,t)^{i-j} ,$$

and in particular

$$P_{(j)}(x) = \frac{(m+j)!}{(m-j)! \, j!} (-1)^j \, H(x,y)^{m-j} .$$

Likewise, we obtain

$$P_{(j)}(y) = (-1)^m \frac{(m+j)!}{(m-j)! \, j!} H(x,y)^{m-j} = (-1)^{m-j} P_{(j)}(x).$$

Now consider the remainder

$$\rho_m = \frac{1}{(2m)!} \int_x^y d_m[H(x,t)H(t,y)]^m f_{(m)}(t)h(t) \, dt = \frac{m!}{(2m)!} \int_x^y P_{(0)}(t)f_{(m)}(t)h(t) \, dt.$$

For $m \geq 1$, repeated integration by parts yields

$$\rho_m = \int_x^y f(t)dt - \frac{m!}{(2m)!} (-1)^m \left[\sum_{j=0}^{m-1} (-1)^j f_{(j)}(t) \, P_{(m-j-1)}(t) \right]_x^y$$

from which we obtain

$$\int_x^y f(t)dt = \frac{m!}{(2m)!} \sum_{j=0}^{m-1} H(x,y)^{j+1} \frac{(2m-j-1)!}{(j+1)!(m-j-1)!} [f_{(j)}(x) + (-1)^j f_{(j)}(y)] + \rho_m.$$

The result follows by first splitting the integral as follows, $\int_a^b f(t) \, dt$

$= \sum_{i=0}^{n-1} \int_{t_{i,n}}^{t_{i+1,n}} f(t) \, dt$, and then applying the above formula in each subinterval

$[t_{i,n}, t_{i+1,n}]$ and using $H(t_{i,n}, t_{i+1,n}) = 1/n$. The second expression is

obtained by m-times repeated integration by parts of the remainder ρ_m. ◻

Based on Proposition 1 we introduce the following rule of degree m using

derivatives of f based on regular sampling $T_n(h)$,

$$_m I_n(f;h) = \frac{m!}{(2m)!} \sum_{j=0}^{m-1} \frac{1}{n^{j+1}} \frac{(2m-j-1)!}{(m-j-1)!(j+1)!} \sum_{i=0}^{n-1} [f_{(j)}(t_{i,n}) + (-1)^j f_{(j)}(t_{i+1,n})]$$

as an approximation to the integral $I(f) = \int_a^b f(t)dt$. For m=1 and m=2, these

rules are identical to the rules used in BC (1989). Also for j: odd the i sum

telescopes and reduces to $f_{(j)}(a)-f_{(j)}(b)$. Their asymptotic performance is as

follows.

Proposition 2. If f and h have p≥m continuous derivates on [a,b], then

$$n^p[I(f) - {}_m I_n(f;h)] \xrightarrow[n]{} 0 \qquad\qquad \text{for } m \leq p < 2m,$$

$$\underset{n}{\to} \frac{(-1)^m m!^2}{(2m)!\,(2m+1)!} [f_{(2m-1)}(b) - f_{(2m-1)}(a)] \text{ for } p \geq 2m.$$

Proof. For simplicity of notation we write t_i for $t_{i,n}$ and I_n for $_mI_n(f;h)$.

If f and h have $(m+j)$, $0 \leq j \leq m$, continuous derivatives, then j-times repeated integration by parts of the remainder ρ_m in Proposition 1 gives

$$\rho_m = \frac{(-1)^j}{(2m)!} \sum_{i=0}^{n-1} \int_{t_i}^{t_{i+1}} d_{m-j} [H(t_i,t)\, H(t,t_{i+1})]^m\, f_{(m+j)}(t) h(t)\, dt.$$

The function $d_{m-j}[H(t_i,t)\,H(t,t_{i+1})]^m$ has $q \leq j$ distinct roots in $[t_i, t_{i+1}]$, say $c_{i,1} < \ldots < c_{i,q}$ so that putting $c_{i,0} = t_i$, $c_{i,q+1} = t_{i+1}$, we have

$$\rho_m = \frac{(-1)^j}{(2m)!} \sum_{i=0}^{n-1} \sum_{\ell=0}^{q} \int_{c_{i,\ell}}^{c_{i,\ell+1}} d_{m-j}[H(t_i,t)H(t,t_{i+1})]^m\, f_{(m+j)}(t) h(t)\, dt.$$

Since $d_{m-j}[H(t_i,t)H(t,t_{i+1})]^m$ has constant sign in each subinterval $(c_{i,\ell}, c_{i,\ell+1})$, the Mean Value Theorem can be applied to obtain

$$\rho_m = \frac{(-1)^j}{(2m)!} \sum_{i=0}^{n-1} \sum_{\ell=0}^{q} f_{(m+j)}(\xi_{i,\ell}) \int_{c_{i,\ell}}^{c_{i,\ell+1}} d_{m-j}[H(t_i,t)H(t,t_{i+1})]^m\, h(t)\, dt$$

where $c_{i,\ell} < \xi_{i,\ell} < c_{i,\ell+1}$. If we write $\int_{t_i}^{c_{i,\ell}} h(t)\, dt = \alpha_\ell/n$, it follows that $0 = \alpha_0 < \alpha_1 < \ldots < \alpha_q < \alpha_{q+1}$ and putting $u = nH(t_i,t)$, we obtain

$$\rho_m = \frac{(-1)^j}{(2m)!} \sum_{i=0}^{n-1} \sum_{\ell=0}^{q} f_{(m+j)}(\xi_{i,\ell}) \frac{1}{n^{m+j+1}} \int_{\alpha_\ell}^{\alpha_{\ell+1}} d^{m-j}[u(1-u)]^m\, du.$$

From the Mean Value Theorem, we have $\frac{1}{n} = h(\zeta_i)(t_{i+1}, t_i)$, $t_i < \zeta_i < t_{i+1}$, so that by Riemann integrability, we have

$$n^{m+j}[I - I_n] \underset{n}{\to} \frac{(-1)^j}{(2m)!} \left\{ \sum_{\ell=0}^{q} \int_{\alpha_\ell}^{\alpha_{\ell+1}} d^{m-j}[u(1-u)]^m du \right\} \int_a^b f_{(m+j)}(t) h(t) dt$$

$$= \frac{(-1)^j}{(2m)!} \int_0^1 d^{m-j}[u(1-u)]^m\, du\, [f_{(m+j-1)}(b) - f_{(m+j-1)}(a)]$$

$$= \frac{(-1)^m}{(2m)!} B(m,m) \left[f_{(2m-1)}(b) - f_{(2m-1)}(a) \right] \quad \text{if} \quad j=m$$

$$= 0 \qquad\qquad\qquad\qquad \text{if} \quad 0 \leq j < m.$$

Using $B(m,m) = \frac{\Gamma(m+1)\ \Gamma(m+1)}{\Gamma(2m+2)} = \frac{m!^2}{(2m+1)!}$, we obtain the final result.

Now assume $j > m$. Then m-times repeated integration by parts of the remainder ρ_m in Proposition 1 yields

$$\rho_m = \frac{(-1)^m}{(2m)!} \sum_{i=0}^{n-1} \int_{t_i}^{t_{i+1}} \left[H(t_i,t)H(t,t_{i+1}) \right]^m f_{(2m)}(t)h(t)\ dt$$

Using the explicit expression of $P_{(0)}(t)$ in the proof of Proposition 1, we have

$$\rho_m = \frac{(-1)^m}{(2m)!} \sum_{\ell=0}^{m} \binom{m}{\ell} (-1)^\ell \frac{1}{n^{m-\ell}} \sum_{i=0}^{n-1} \int_{t_i}^{t_{i+1}} H(t_i,t)^{m+\ell} f_{(2m)}(t)h(t)\ dt.$$

Again, repeated integration by parts gives

$$\int_{t_i}^{t_{i+1}} H(t_i,t)^{m+\ell} f_{(2m)}(t)h(t)\ dt$$

$$= \sum_{p=0}^{j-m-1} \left[\frac{H(t_i,t)^{m+\ell+1+p}}{(m+\ell+1)\ldots(m+\ell+1+p)} f_{(2m+p)}(t) \right]_{t_i}^{t_{i+1}}.$$

$$+ (-1)^{j-m} \int_{t_i}^{t_{i+1}} \frac{1}{(m+\ell+1)\ldots(j+\ell)} H(t_i,t)^{j+\ell} f_{(j+m)}(t)h(t)\ dt$$

$$= \sum_{p=0}^{j-m-1} \frac{(m+\ell)!}{(m+\ell+1+p)!} \frac{1}{n^{m+\ell+1+p}} f_{(2m+p)}(t_{i+1})$$

$$+ (-1)^{j-m} \frac{(j+\ell)!}{(m+\ell)!} \int_{t_i}^{t_{i+1}} H^{j+\ell}(t_i,t) f_{(j+m)}(t)h(t)\ dt$$

and using the Mean Value Theorem, we have, with $t_i < \xi_i < t_{i+1}$,

$$= \frac{1}{(m+\ell+1)n^{m+\ell+1}} f_{(2m)}(t_{i+1}) + \sum_{p=1}^{j-m-1} \frac{(m+\ell)!}{(m+\ell+1+p)!} \frac{1}{n^{m+\ell+1+p}} f_{(2m+p)}(t_{i+1})$$

$$+ (-1)^{j-m} \frac{(j+\ell)!}{(m+\ell)!} \frac{1}{(j+\ell+1)n^{j+\ell+1}} f_{(m+j)}(\xi_i)$$

$$= \frac{1}{(m+\ell+1)n^{m+\ell+1}} f_{(2m)}(t_{i+1}) + o(n^{-(m+\ell+1)}).$$

Therefore we obtain

$$\rho_m = \frac{(-1)^m}{(2m)!} \sum_{\ell=0}^{m} \binom{m}{\ell}(-1)^{\ell} \frac{1}{m+\ell+1} \frac{1}{n^{2m+1}} \sum_{i=0}^{n-1} f_{(2m)}(t_{i+1}) + o(n^{-2m})$$

$$= \frac{(-1)^m}{(2m)!} \int_0^1 [u(1-u)]^m \, du \, \frac{1}{n^{2m+1}} \sum_{i=0}^{n-1} f_{(2m)}(t_{i+1}) + o(n^{-2m}).$$

Using $\frac{1}{n} = h(\zeta_i) (t_{i+1}-t_i)$, we obtain

$$n^{2m}[I-I_n] \underset{n}{\rightarrow} \frac{(-1)^m}{(2m)!} B(m,m) \int_a^b f_{(2m)}(t)h(t) \, dt$$

$$= \frac{(-1)^m m!^2}{(2m)!(2m+1)!} [f_{(2m-1)}(b) - f_{(2m-1)}(a)] . \qquad \square$$

3. OPTIMAL-COEFFICIENT ESTIMATORS

The optimal-coefficient estimators $_d\hat{I}_n$ minimize the mean square error $E(I-_dI_n)^2$ among all linear estimators that use the q.m. derivatives of the process X at the sampling points. They are of the form

(3.1) $$_d\hat{I}_n = f'_{K,T_n} R^{-1}_{K,T_n} Y_{T_n}$$

where

$$Y'_{T_n} = (X^{(j)}(t_{0,n}), \ldots, X^{(j)}(t_{n,n}))_{j=0,\ldots,K} : 1 \times (K+1)(n+1) \text{ vector},$$

$$R_{K,T_n} = \{R^{(\ell,m)}(t_{i,n}, t_{j,n})\}_{\ell,m=0,\ldots,K}^{i,j=0,\ldots,n} : (K+1)(n+1) \times (K+1)(n+1)$$

covariance matrix assumed nonsingular,

$$f'_{K,T_n} = (f^{(j)}(t_{0,n}), \ldots, f^{(j)}(t_{n,n}))_{j=0,\ldots,K} : 1 \times (K+1)(n+1) \text{ vector},$$

$$f(t) = \int_a^b R(s,t) \phi(s)ds.$$

The asymptotic performance of these optimal-coefficient estimators using regular sequences of sampling designs, $_d\hat{I}_n(h)$, is given by Theorem 1 under slightly different assumptions than those considered by Sacks and Ylvisaker, where also the jump function $\alpha_K(\cdot)$ was assumed to be constant.

<u>Theorem 1.</u> Under Assumption A_K we have

$$n^{2K+2} E[I - _d\hat{I}_n(h)]^2 \underset{n}{\rightarrow} \frac{(K+1)!^2}{(2K+2)!(2K+3)!} \int_a^b \frac{\alpha_K(t)\phi^2(t)}{h^{2K+2}(t)} dt.$$

If h^* is proportional to $(\alpha_K\phi^2)^{1/(2K+3)}$, then the asymptotic performance becomes

$$n^{2K+2} E[I - _d\hat{I}_n(h^*)]^2 \underset{n}{\rightarrow} \frac{(K+1)!^2}{(2K+2)!(2K+3)!} \{\int_a^b [\alpha_K(t)\phi^2(t)]^{1/(2K+3)} dt\}^{2K+3}$$

with minimal value of the asymptotic constant. The asymptotic optimality of the designs $\{T_n(h^*)\}_n$ is shown in Sacks and Ylvisaker (1970), where the jump function α_K was assumed constant, and under a Lipschitz type assumption near the zeroes of weight function ϕ (cf. (3.15) in SY (1970b)).

The condition A_K and function α_K used in Theorem 1 are defined as follows, where the following notations are used: $R^{(p,q)}(s,t) = \partial^{p+q}R(s,t)/\partial s^p\partial t^q$, $R^{(p,q)}(t,t-) = \lim_{s\uparrow t} R^{(p,q)}(t,s)$ and $R^{(p,q)}(t,t+) = \lim_{s\downarrow t} R^{(p,q)}(t,s)$.

<u>Assumption</u> A_K (K=0,1,2,...).(i) $R^{(K,K)}$ exist and are continuous on the square $[a,b] \times [a,b]$.

(ii) Each $R^{(p,q)}(t,s)$ with $p+q=2K$ exists on the square $[a,b]\times[a,b]$; has continuous mixed partial derivatives up to order 2 off the diagonal $(t\neq s)$; and at the diagonal $(t=s)$ it has left and right derivatives which are continuous functions of t, i.e. $R^{(p,q)}(t+,t)$, $R^{(p,q)}(t-,t)$ exist and are finite for all $p+q = 2K+1$ and are continuous functions of t, and $\sup_{s\neq t} |R^{(0,2K+1)}(s,t)| < \infty$. For each $t \in [a,b]$, $R^{(0,2K+2)}(\cdot,t+) \in H(R)$, the reproducing kernel Hilbert

space of the covariance R with norm $\|\cdot\|_R$, and $\sup_t \|R^{(0,2K+2)}(\cdot,t+)\|_R < \infty$. Also each $R^{(j,2K+2)}(t,s)$, $j=1,\ldots,K$, exist and are continuous on $[a,b]\times[a,b]$ for $t \neq s$.

(iii) $\alpha_K(t) = R^{(K,K+1)}(t,t-) - R^{(K,K+1)}(t,t+)$ is positive and continuous on $[a,b]$.

(iv) ϕ and h have K+2 continuous derivatives on $[a,b]$.

Part (i) of Assumption A_K is the necessary and sufficient condition for the process X to have K mean-square continuous q.m. derivatives. Part (ii) requires smoothness properties off the diagonal and thus it is weak. Part (iii) guarantees the process X has no more than K q.m. derivatives. Assumptions (i)-(iii) are satisfied by a large class of processes including K^{th} order iterated integrals of Wiener process, and stationary processes with rational spectral densities. When X is stationary, $R(t,s) = R(t-s)$, then conditions (i)-(iii) are satisfied iff $R^{(2K+2)}(t)$ exists and is continuous for $t \neq 0$ and $R^{(2K+1)}(0+)$, $R^{(2K+1)}(0-)$ exist, are finite and the jump $\alpha_K(t) = R^{(2K+1)}(0-) - R^{(2K+1)}(0+) = \alpha_K$ is positive.

In the proof of Theorem 1, we will use the following "h-weighted" derivatives of the covariance R,

$$R_{(p,q)}(t,s) = E\{Y_{(p)}(t)Y_{(q)}(s)\} \qquad \text{for} \qquad 0 \leq p,q \leq K, \text{ and } t,s \in [a,b].$$

Furthermore, we define recursively $R_{(p,q)}(t,s)$ for $\{K < p$ or $K < q$ and $0 \leq p+q \leq 2K$, $t,s \in [a,b]\}$ and for $\{2K+1 \leq p+q \leq 2K+2$, $t \neq s$ in $[a,b]\}$ as follows:

$$R_{(p,q)}(t,s) = \frac{1}{h(s)} R^{(0,1)}_{(p,q-1)}(t,s) \qquad \text{for } K < q,$$

$$R_{(p,q)}(t,s) = \frac{1}{h(t)} R^{(1,0)}_{(p-1,q)}(t,s) \qquad \text{for } K < p.$$

<u>Proof of Theorem 1</u>. Let $P_{K,n}$ be the projection from the Hilbert space $H(R_{(0)})$

onto $\text{sp}\{R_{(0,j)}(..t), \ t \in T_n, \ j = 0,\ldots,K\}$. In view of the isomorphism between $H(R_{(0)})$ and $H(Y_{(0)})$, as in the proof of Theorem 1 in BC (1989), the mean square error of the estimation of I by $_d\hat{I}_n(h)$ can be written as follows:

$$E[I-_d\hat{I}_n(h)]^2 = \|f-P_{K,n}f\|^2_{R_{(0)}} = \langle f-P_{K,n}f, \ f\rangle_{R_{(0)}} = \int_a^b (f-P_{K,n}f)(s)\ h(s)\ ds$$

where $f(t) = \int_a^b R_{(0)}(t,s)\ h(s)\ ds$. Let $g = f - P_{K,n}f$. Then g is orthogonal to $\text{sp}\{R_{(0,j)}(\cdot,t), \ t \in T_n, \ j = 0,\ldots,K\}$ and in particular $g_{(j)}(t_i) = \langle g, \ R_{(0,j)}(\cdot,t_i)\rangle_{R_{(0)}} = 0$ for $j = 0,\ldots,K$. Applying Proposition 1 with $m = K+1$ to the function g, we obtain

$$E[I-_d\hat{I}_n(h)]^2 = \int_a^b g(s)\ h(s)\ ds$$

$$= \frac{(K+1)!}{(2K+2)!} \sum_{i=0}^{n-1} \sum_{j=0}^{K} \frac{1}{n^{j+1}} \frac{(2K+1-j)!}{(K-j)!\ (j+1)!} [g_{(j)}(t_i) + (-1)^j g_{(j)}(t_{i+1})]$$

$$+ \frac{(-1)^{K+1}}{(2K+2)!} \sum_{i=0}^{n-1} \int_{t_{i,n}}^{t_{i+1,n}} [H(t_{i,n},x)H(x,t_{i+1,n})]^{K+1}\ g_{(2K+2)}(x)h(x)\ dx$$

$$= \frac{(-1)^{K+1}}{(2K+2)!} \sum_{i=0}^{n-1} \int_{t_{i,n}}^{t_{i+1,n}} [H(t_{i,n},x)H(x,t_{i+1,n})]^{K+1}\ g_{(2K+2)}(x)h(x)\ dx.$$

From Assumption A_K, we have

$$f_{(2K+1)}(t) = \int_a^t R_{(0,2K+1)}(s,t+)h(s)ds + \int_t^b R_{(0,2K+1)}(s,t-)h(s)ds$$

and taking the h-weighted derivative,

$$f_{(2K+2)}(t) = -\beta_K(t) + \int_a^b R_{(0,2K+2)}(s,t+)h(s)ds$$

where $\beta_K(t) = R_{(0,2K+1)}(t,t-) - R_{(0,2K+1)}(t,t+)$. From Assumption A_K, $R_{(0,j)}(\cdot,s) \in H(R_{(0)})$ for $j=0,\ldots,K$ and $R_{(0,2K+2)}(\cdot,t+) \in H(R_{(0)})$, and thus

$$\langle R_{(0,j)}(\cdot,s), \ R_{(0,2K+2)}(\cdot,t+)\rangle_{R_{(0)}} = R_{(j,2K+2)}(s,t+).$$

and since $P_{K,n}f \in \text{sp}\{R_{(0,j)}(\cdot,t), t \in T_n, \ j=0,\ldots,K\}$, we have for $t \notin T_n$,

$$\langle P_{K,n}f, \ R_{(0 \ 2K+2)}(\cdot,t+)\rangle_{R_{(0)}} = (P_{K,n}f)_{(2K+2)}(t+) = (P_{K,n}f)_{(2K+2)}(t).$$

Writing $R_{(0,2K+2)}(\cdot,t+) = EY_{(0)}\xi_t$, $\xi_t \in H(R_{(0)})$, it follows from the isomorphism between $H(R_{(0)})$ and $H(Y_{(0)})$ that

$$\langle f, R_{(0,2K+2)}(\cdot,t+)\rangle_{R_{(0)}} = E[\int_a^b Y_{(0)}(s)h(s)ds \cdot \xi_t]$$

$$= \int_a^b E[Y_{(0)}(s)\xi_t]h(s)ds = \int_a^b R_{(0,2K+2)}(s,t+)h(s).$$

It then follows that for $t \notin T_n$,

$$g_{(2K+2)}(t) = -\beta_K(t) + \langle f - P_{K,n}f, \ R_{(0,2K+2)}(\cdot,t+) \rangle_{R_{(0)}}.$$

Then, it follows that

$$E[I - {}_d\hat{I}_n(h)]^2 = \frac{(-1)^K}{(2K+2)!} \sum_{i=0}^{n-1} \int_{t_{i,n}}^{t_{i+1,n}} \beta_K(t)[H(t_{i,n},t)H(t,t_{i+1,n})]^{K+1} h(t) \, dt$$

$$+ \frac{(-1)^{K+1}}{(2K+2)!} \sum_{i=0}^{n-1} \int_{t_{i,n}}^{t_{i+1,n}} \langle f - P_{K,n}f, \ R_{(0,2K+2)}(\cdot,t+) \rangle_{R_{(0)}}$$

$$\times [H(t_{i,n},t)H(t,t_{i+1,n})]^{K+1}h(t) \, dt.$$

<u>The second term</u> is bounded in absolute value by:

$$\frac{1}{(2K+2)!} \| f - P_{K,n}f\|_{R_{(0)}} M_{0,2K+2} \sum_{i=0}^{n-1} \int_{t_{i,n}}^{t_{i+1,n}} [H(t_{i,n},t)H(t,t_{i+1,n})]^{K+1}h(t)dt,$$

where $M_{0,2K+2} = \sup_t \|R_{(0,2K+2)}(\cdot,t+)\|_{R_{(0)}} < \infty$. As in the proof of Proposition 2, we have

$$\sum_{i=0}^{n-1} \int_{t_{i,n}}^{t_{i+1,n}} [H(t_{i,n},t)H(t,t_{i+1,n})]^{K+1} h(t) \, dt = \frac{(K+1)!^2}{n^{2K+2} (2K+3)!}.$$

The upper bound then becomes (for some constant c_0).

$$c_0 \; \|f - P_{K,n} f\|_{R_{(0)}} \; n^{-(2K+2)}.$$

Since $\|\beta_K(t)\| \leq c_1$ by the continuity of $\beta_K(t)$ on $[a,b]$, and $\|f - P_{K,n} f\|_{R_{(0)}} \leq \|f\|_{R_{(0)}}$, we have for $t \in T_n$,

$$|g_{(2K+2)}(t)| \leq c_1 + \|f\|_{R_{(0)}} M_{0,2K+2}.$$

so that for some constant c_2,

$$\|f - P_{K,n} f\|^2_{R_{(0)}} = \frac{(-1)^{K+1}}{(2K+2)!} \sum_{i=0}^{n-1} \int_{t_{i,n}}^{t_{i+1,n}} [H(t_{i,n},x) H(x, t_{i+1,n})]^{K+1} \; g_{(2K+2)}(x) h(x) \; dx$$

$$\leq c_2 n^{-(2K+2)}.$$

It follows that for some constant c_3 the second term is upper bounded by $c_3 n^{-(3K+3)}$. Therefore the second term is $o(n^{-(3K+2)})$.

For <u>the first term</u>, since each $H(t_{i,n}, t) H(t, t_{i+1,n})$ has positive sign on $(t_{i,n}, t_{i+1,n})$, the Mean Value Theorem can be applied to write it in the form

$$\frac{(-1)^K}{(2K+2)! n^{2K+2}} \frac{(K+1)!^2}{(2K+3)!} \sum_{i=0}^{n-1} \beta_K(\xi_{i,n}) \frac{1}{n}$$

where $t_{i,n} < \xi_{i,n} < t_{i+1,n}$. Using $\frac{1}{n} = h(\zeta_{i,n})(t_{i+1,n} - t_{i,n})$, $t_{i,n} < \zeta_{i,n} < t_{i+1,n}$, and the above result for the first term, we obtain by Riemann integrability,

$$n^{2K+2} E[I - {}_d\hat{I}_n(h)]^2 \to \frac{(-1)^K (K+1)!^2}{(2K+2)!(2K+3)!} \int_a^b \beta_K(t) h(t) \; dt.$$

The final expression of the asymptotic constant follows from Lemma 3 in BC (1989). □

4. SIMPLE-COEFFICIENT ESTIMATORS

Simple-coefficient estimators using the existing K q.m. derivatives of X at the sampling points give better approximation of the integral I than the

simple-coefficient estimators (1.3) which do not use these q.m. derivatives. These estimators are based on the trapezoidal rule, with a correction term that depends on the values of the q.m. derivatives of X at all sampling points of a regular sequence of sampling designs, i.e. on the weighted Rodriguez formula. They are of the form:

$$(4.1) \quad {}_d I_n(h) = \frac{(K+1)!}{(2K+2)!} \sum_{j=0}^{K} \frac{1}{n^{j+1}} \frac{(2K+1-j)!}{(j+1)!(K-j)!} \sum_{i=0}^{n-1} [Y_{(j)}(t_{i,n}) + (-1)^j Y_{(j)}(t_{i+1,n})],$$

where the h-weighted q.m. derivatives $Y_{(j)}$ of Y are defined as in Section 2:

$$Y_0 = Y/h, \quad Y_{(j)} = Y^{(1)}_{(j-1)}/h, \quad j \geq 1,$$

and superscript denotes q.m. derivative. Note that for odd j the i-sum telescopes to $Y_{(j)}(a) - Y_{(j)}(b)$. Their asymptotic performance is given by Theorem 2 under a weaker assumption than in Theorem 1.

Theorem 2. Under Assumption A'_K, we have

$$n^{2K+2} E[I - {}_d I_n(h)]^2 \xrightarrow[n]{} \frac{(K+1)!^2}{(2K+2)!(2K+3)!} \int_a^b \frac{\phi^2(t)\alpha_K(t)}{h^{2K+2}(t)} dt.$$

If h^* is proportional to $(\alpha_K \phi^2)^{1/(2K+3)}$, then the remarks on the asymptotic performance and optimality following Theorem 1 are also applicable here. Assumption A'_K is defined as follows, and the comments made on Assumption A_K in Section 3 are applicable.

Assumption A'_K (K=0,1,2,...). Assume (i), (iii), (iv) of A_K and instead of (ii) the following.

(ii)' If p,q are nonnegative integers with p+q≤2, then $R^{(K+p,K+q)}(s,t)$ exists and is continuous off the diagonal of the square $[a,b] \times [a,b]$ (i.e. for s≠t); at the diagonal s=t, $R^{(K,K+1)}(t,t+)$, $R^{(K,K+1)}(t,t-)$ exist, are finite and are continuous functions of t; and for p,q≤1, $\sup_{s \neq t} |R^{(K+p,K+q)}(s,t)| < \infty$.

For K=0, the simple-coefficient estimators have only a "trapezoidal" component (without correction term). For K=1, they have a correction term that depends only on the q.m. derivatives of X at the endpoints of the interval [a,b], and after approximating these q.m. derivatives by finite differences, the resulting estimators are identical to those given in (1.3).

The weighted q.m. derivatives in (4.1) can be approximated by the generalized version of Newton's backward and forward finite difference formulas for regular sampling as follows (cf. BC (1989), Krylov (1962))

$$Y_{(j)}(t_{i,n}) \approx n^j \sum_{\ell=j}^{K} \frac{1}{\ell!} W_\ell^{(j)}(0) \Delta^\ell Y_{(0)}(t_{i,n}) \qquad \text{for } i=0,1,\ldots,n-K,$$

$$\approx n^j \sum_{\ell=j}^{K} \frac{(-1)^{\ell+j}}{\ell!} W_\ell^{(j)}(0) \Delta^\ell Y_{(0)}(t_{i-\ell,n}) \qquad \text{for } i=n-K+1,\ldots,n,$$

where $W_\ell(u)=u(u-1)\ldots(u-\ell+1)$, $\ell \geq 1$ ($W_0(u)=1$). In order for the procedure to be "symmetric", we also use in (4.1),

$$Y_{(j)}(t_{i+1,n}) \approx n^j \sum_{\ell=j}^{K} \frac{1}{\ell!} W_\ell^{(j)}(0) \Delta^\ell Y_{(0)}(t_{i+1,n}) \qquad \text{for } i=0,\ldots,K-2,$$

$$\approx n^j \sum_{\ell=j}^{K} \frac{1}{\ell!} W_\ell^{(j)}(0) \Delta^\ell Y_{(0)}(t_{i+1-\ell,n}) \qquad \text{for } i=K-1,\ldots,n-1.$$

The estimators derived from these approximations and (4.1) are of the form

$$(4.2) \qquad d\overline{I}_n(h) = \frac{1}{n} \sum_{i=0}^{n} b_i \left(\frac{\phi X}{h}\right)(t_{i,n})$$

where the expressions of the coefficients b_i are too complicated to write for general K (much more so than in (1.4")). Their values for K=0,1,2,3 (and appropriately large n) are:

$$K = 0: \quad \tfrac{1}{2}, \quad 1, \quad 1, \quad 1, \ 1, \ 1, \ 1, \ldots, 1, \ 1, \ 1, \quad 1, \quad 1, \quad 1, \quad \tfrac{1}{2}$$

$$K = 1: \quad \tfrac{5}{12}, \quad \tfrac{13}{12}, \quad 1, \quad 1, \ 1, \ 1, \ 1, \ldots, 1, \ 1, \ 1, \quad 1, \quad 1, \quad \tfrac{13}{12}, \quad \tfrac{5}{12}$$

$$K = 2: \quad \tfrac{11}{30}, \quad \tfrac{143}{120}, \quad \tfrac{14}{15}, \quad \tfrac{121}{120}, \ 1, \ 1, \ 1, \ldots, 1, \ 1, \ 1, \quad \tfrac{121}{120}, \quad \tfrac{14}{15}, \quad \tfrac{143}{120}, \quad \tfrac{11}{30}$$

$$K = 3: \tfrac{529}{1680}, \tfrac{2263}{1680}, \tfrac{1327}{1680}, \tfrac{1721}{1680}, \tfrac{87}{84}, \tfrac{83}{84}, \ 1, \ldots, 1, \ \tfrac{83}{84}, \tfrac{87}{84}, \tfrac{1721}{1680}, \tfrac{1327}{1680}, \tfrac{2263}{1680}, \tfrac{529}{1680}$$

We conjecture that the simple-coefficient estimators (4.2) have the same asymptotic performance (1.1) as the simple-coefficient estimators (1.3). The example in Section 5 of a stationary process with two q.m. derivatives seems to support this claim.

The asymptotic constant corresponding to the optimal-coefficient estimators or the simple-coefficient estimators (1.3) that do not use the existing K q.m. derivatives is

$$C_K = \frac{|B_{2K+2}|}{(2K+2)!} \int_a^b \frac{\alpha_K(t)\phi^2(t)}{h^{2K+2}(t)} \, dt.$$

The asymptotic constant corresponding to the optimal-coefficient estimators or the simple-coefficient estimators (4.1) that use the existing K q.m. derivatives is:

$$_d C_K = \frac{(K+1)!^2}{(2K+2)!(2K+3)!} \int_a^b \frac{\alpha_K(t)\phi^2(t)}{h^{2K+2}(t)} \, dt.$$

Their ratio depends only on the degree of smoothness K:

$$\frac{_d C_K}{C_K} = \frac{(K+1)!^2}{|B_{2K+2}|(2K+3)!} \ .$$

For large sample sizes, the number of samples required for a given mean square error when using estimators with and without q.m. derivatives are related by:

$$\frac{d^n_K}{n_K} = \left[\frac{d^C_K}{C_K}\right]^{1/(2K+2)} = \begin{cases} 1 & \text{for } K = 0,1, \\ .82 & \text{for } K = 2, \\ .60 & \text{for } K = 3, \\ .41 & \text{for } K = 4. \end{cases}$$

Thus, the performance improvement of the estimators that use q.m. derivatives over those that do not increases as the number of q.m. derivatives K increases.

<u>Proof of Theorem 2</u>.   For simplicity we write I_n for $_d I_n(h)$.   In view of the isomorphism between $H(R_{(0)})$ and $H(Y_{(0)})$, the mean square error is written as follows

$$E(I - I_n)^2 = \|f_n\|^2_{R_{(0)}}$$

where

$$f_n(t) = E[Y_{(0)}(t)(I-I_n)]$$

$$= \int_a^b R_{(0)}(t,s)h(s)ds - \sum_{i=0}^{n-1}\sum_{j=0}^{K}\frac{1}{n^{j+1}} F_{j,K}[R_{(0,j)}(t,t_i)+(-1)^j R_{(0,j)}(t,t_{i+1})]$$

and

$$F_{j,K} \equiv \frac{(K+1)!}{(2K+2)!} \frac{(2K-j+1)!}{(K-j)!(j+1)!} .$$

Thus, for $K \geq 0$,

$$E(I-I_n)^2 = \langle f_n, f_n \rangle_{R_{(0)}}$$

$$= \int_a^b f_n(t)h(t)dt - \sum_{i=0}^{n-1}\sum_{j=0}^{K}\frac{1}{n^{j+1}} F_{j,K} [(f_n)_{(j)}(t_i) + (-1)^j (f_n)_{(j)}(t_{i+1})]$$

$$= \frac{1}{(2K+2)!} \sum_{i=0}^{n-1} \int_{t_i}^{t_{i+1}} d_{K+1}[H(t_i,t)H(t,t_{i+1})]^{K+1}(f_n)_{(K+1)}(t)h(t)dt$$

by Proposition 1 with $m = K+1$.   Integrating by parts we obtain

$$E(I-I_n)^2 = \frac{1}{(2K+2)!} \sum_{i=0}^{n-1} \{ [d_{K+1}[H(t_i,t)H(t,t_{i+1})]^{K+1}(f_n)_{(K)}(t)]_{t_i}^{t_{i+1}}$$

$$- \int_{t_i}^{t_{i+1}} d_{K+2}[H(t_i,t)H(t,t_{i+1})]^{K+1}(f_n)_{(K)}(t)h(t)dt \ \}.$$

Applying Proposition 1 to the integral of $R_{(0)}(t,\cdot)$ in the expression of f_n, we obtain

$$f_n(t) = \frac{1}{(2K+2)!} \sum_{i=0}^{n-1} \int_{t_i}^{t_{i+1}} d_{K+1}[H(t_i,s)H(s,t_{i+1})]^{K+1} R_{(0,K+1)}(t,s)h(s)ds,$$

and taking the Kth weighted derivative, we have

$$(f_n)_{(K)}(t) = \frac{1}{(2K+2)!} \sum_{j=0}^{n-1} \int_{t_j}^{t_{j+1}} d_{K+1}[H(t_j,s)H(s,t_{j+1})]^{K+1} R_{(K,K+1)}(t,s)h(s)ds.$$

Now we show that $(f_n)_{(K)}(t_i) = (f_n)_{(K)}(t_{i+1}) = 0$. Indeed, using the following function which was introduced in the proof of Proposition 1,

$$P_{(0)}(t;t_i,t_{i+1}) = \frac{1}{(K+1)!} d_{K+1}[H(t_i,t)H(t,t_{i+1})]^{K+1},$$

and its explicit expression there, we have

$$(f_n)_{(K)}(t_i) = \frac{(K+1)!}{(2K+2)!} \sum_{j=0}^{n-1} \int_{t_j}^{t_{j+1}} P_{(0)}(s;t_j,t_{j+1})R_{(K,K+1)}(t_i,s)h(s) \ ds$$

$$= \frac{(K+1)!}{(2K+2)!} \sum_{\ell=0}^{K+1} \frac{(K+1+\ell)!(-1)^\ell}{(K+1-\ell)! \ \ell!^2} \frac{1}{n^{K+1-\ell}} \sum_{j=0}^{n-1} \int_{t_j}^{t_{j+1}} H^\ell(t_j,s)R_{(K,K+1)}(t_i,s)h(s) \ ds.$$

Applying the Mean Value Theorem, we have

$$(f_n)_{(K)}(t_i) = \frac{(K+1)!}{(2K+2)!} \frac{1}{n^{K+2}} \sum_{\ell=0}^{K+1} \frac{(K+1+\ell)!(-1)^\ell}{(K+1-\ell)! \ \ell!^2(\ell+1)} \sum_{j=0}^{n-1} R_{(K,K+1)}(t_i,b_j)$$

where $t_j < b_j < t_{j+1}$. By taking h to be the uniform density, the h-weighted polynomial $P_{(0)}(t;t_i,t_{i+1})$ becomes the classical Legendre polynomial $\frac{1}{(K+1)!} D^{K+1}[u(1-u)]^{K+1}$ over the interval $[0,1]$ and therefore

$$\int_0^1 D^{K+1}[u(1-u)]^{K+1}du = \sum_{\ell=0}^{K} \frac{(K+1+\ell)! \ (-1)^\ell}{(K+1-\ell)! \ \ell!^2} \int_0^1 u^\ell \ du = \sum_{\ell=0}^{K} \frac{(K+1+\ell)! \ (-1)^\ell}{(K+1-\ell)! \ \ell!^2(\ell+1)}.$$

and since

$$\int_0^1 D^{K+1}[u(1-u)]^{K+1} \, du = \{D^K[u(1-u)]^{K+1}\}_0^1 = 0,$$

it follows that $(f_n)_{(K)}(t_i) = 0$, and likewise $(f_n)_{(K)}(t_{i+1}) = 0$. Thus the first term in the expression of the m.s.e. $E(I-I_n)^2$ vanishes and we can write

$$E(I-I_n)^2 = - \frac{(K+1)!}{(2K+2)!} \sum_{i=0}^{n-1} \int_{t_i}^{t_{i+1}} P_{(1)}(t;t_i,t_{i+1})(f_n)_{(K)}(t)h(t)dt.$$

Replacing $(f_n)_{(K)}(t)$ by its expression, the m.s.e. is decomposed into two terms: the diagonal terms $\sum_i M_{i,i}$ and the off-diagonal terms $\sum_{i \neq j} M_{i,j}$.

<u>Diagonal terms</u> $M_{i,i}$.

$$M_{i,i} = - \frac{(K+1)!^2}{(2K+2)!^2} \int_{t_i}^{t_{i+1}} P_{(0)}(s;t_i,t_{i+1}) \left[\int_{t_i}^s + \int_s^{t_{i+1}} \right] P_{(1)}(t;t_i,t_{i+1})$$

$$\times R_{(K,K+1)}(t,s)h(t)h(s) \, dtds.$$

Using the expression of $P_{(1)}(t;t_i,t_{i+1})$ in the proof of Proposition 1, and integrating, we obtain from the Mean Value Theorem

$$M_{i,i} = - \frac{(K+1)!}{(2K+2)!^2} \sum_{\ell=1}^{K+1} \frac{(K+1+\ell)!(-1)^\ell}{(K+1-\ell)! \, \ell!(\ell-1)!} \frac{1}{n^{K+1-\ell}}$$

$$\times \left\{ \int_{t_i}^{t_{i+1}} P_{(0)}(s;t_i,t_{i+1}) \frac{1}{\ell} H^\ell(t_i,s)R_{(K,K+1)}(\eta_s,s)h(s) \, ds \right.$$

$$\left. + \int_{t_i}^{t_{i+1}} P_{(0)}(s;t_i,t_{i+1}) \frac{1}{\ell} \left[\frac{1}{n^\ell} - H^\ell(t_i,s) \right] R_{(K,K+1)}(\eta_s',s)h(s) \, ds \right\}$$

where $t_i < \eta_s < s < \eta_s' < t_{i+1}$. From the expression of $P_{(0)}(s;t_i,t_{i+1})$ and the Mean Value Theorem, we have

$$M_{i,i} = - \frac{(K+1)!^2}{(2K+2)!^2} \sum_{\ell=1}^{K+1} \frac{(K+1+\ell)!}{(K+1-\ell)! \, \ell!^2} (-1)^\ell \frac{1}{n^{K+1-\ell}} \sum_{p=0}^{K+1} \frac{(K+1+p)!(-1)^p}{(K+1-p)!p!^2} \frac{1}{n^{K+1-p}}$$

$$\times \left\{ R_{(K,K+1)}(\eta_i,a_i) \int_{t_i}^{t_{i+1}} H^{p+\ell}(t_i,s)h(s) \, ds \right.$$

$$+ R_{(K,K+1)}(\eta_i',a_i') \int_{t_i}^{t_{i+1}} \left[\frac{1}{n^\ell} - H^\ell(t_i,s) \right] H^p(t_i,s) h(s) \, ds \Bigg\}$$

where $t_i < \eta_i < a_i < t_{i+1}$, $t_i < a_i' < \eta_i' < t_{i+1}$.

$$= - \frac{(K+1)!^2}{(2K+2)!^2} \frac{1}{n^{2K+3}}$$

$$\times \Bigg\{ R_{(K,K+1)}(\eta_i,a_i) \sum_{\ell=1}^{K+1} \frac{(K+1+\ell)!}{(K+1-\ell)!} \frac{(-1)^\ell}{\ell!^2} \sum_{p=0}^{K+1} \frac{(K+1+p)!}{(K+1-p)!} \frac{(-1)^p}{p!^2} \frac{1}{p+\ell+1}$$

$$+ R_{(K,K+1)}(\eta_i',a_i') \sum_{\ell=1}^{K+1} \frac{(K+1+\ell)!}{(K+1-\ell)!} \frac{(-1)^\ell}{\ell!^2} \sum_{p=0}^{K+1} \frac{(K+1+p)!}{(K+1-p)!} \frac{(-1)^p}{p!^2} \left[\frac{1}{p+1} - \frac{1}{p+\ell+1} \right] \Bigg\}.$$

We now concentrate on the coefficients of $R_{(K,K+1)}(\eta_i,a_i)$ and $R_{(K,K+1)}(\eta_i',a_i')$.
We have

$$\int_0^1 D^{K+1}[u(1-u)]^{K+1} \, du \int_0^u D^{K+2}[v(1-v)]^{K+1} \, dv$$

$$= (K+1)! \sum_{\ell=1}^{K+1} \frac{(K+1+\ell)!}{(K+1-\ell)!} \frac{(-1)^\ell}{\ell!} \int_0^1 D^{K+1}[u(1-u)]^{K+1} \, du \int_0^u \frac{v^{\ell-1}}{(\ell-1)!} \, dv$$

$$= (K+1)!^2 \sum_{\ell=1}^{K+1} \frac{(K+1+\ell)!}{(K+1-\ell)!} \frac{(-1)^\ell}{\ell!^2} \sum_{p=0}^{K+1} \frac{(K+1+p)!}{(K+1-p)!} \frac{(-1)^p}{p!^2} \int_0^1 u^{p+\ell} \, du$$

$$= (K+1)! \sum_{\ell=1}^{K+1} \frac{(K+1+\ell)!}{(K+1-\ell)!} \frac{(-1)^\ell}{\ell!^2} \sum_{p=0}^{K+1} \frac{(K+1+p)!}{(K+1-p)!} \frac{(-1)^p}{p!^2} \frac{1}{p+\ell+1} .$$

and also by repeated integration by parts we have

$$\int_0^1 D^{K+1}[u(1-u)]^{K+1} \, du \int_0^u D^{K+2}[v(1-v)]^{K+1} \, dv$$

$$= (-1)^{K+1} \int_0^1 [u(1-u)]^{K+1} D^{2K+2}[u(1-u)]^{K+1} \, du$$

$$= (-1)^{K+1} (-1)^{K+1} (2K+2)! \, B(K+1,K+1) = \frac{(2K+2)! \, (K+1)!^2}{(2K+3)!} .$$

Equating the two right hand sides, we find that the coefficient of

$R_{(K,K+1)}(\eta_i,a_i)$ is

$$- \frac{(K+1)!^2}{(2K+2)!(2K+3)!} \frac{1}{n^{2K+3}} .$$

Also,

$$\int_0^1 D^{K+1}[u(1-u)]^{K+1} du \int_u^1 D^{K+2} [v(1-v)]^{K+1} dv$$

$$= (K+1)! \sum_{\ell=1}^{K+1} \frac{(K+1+\ell)!}{(K+1-\ell)!} \frac{(-1)^\ell}{\ell!} \int_0^1 D^{K+1} [u(1-u)]^{K+1} du \int_u^1 \frac{v^{\ell-1}}{(\ell-1)!} dv$$

$$= (K+1)!^2 \sum_{\ell=1}^{K+1} \frac{(K+1+\ell)!}{(K+1-\ell)!} \frac{(-1)^\ell}{\ell!^2} \sum_{p=0}^{K+1} \frac{(K+1+p)!}{(K+1-p)!} \frac{(-1)^p}{p!^2} \int_0^1 u^p (1-u^\ell) du$$

$$= (K+1)!^2 \sum_{\ell=1}^{K+1} \frac{(K+1+\ell)!}{(K+1-\ell)!} \frac{(-1)^\ell}{\ell!^2} \sum_{p=0}^{K+1} \frac{(K+1+p)!}{(K+1-p)!} \frac{(-1)^p}{p!^2} \left[\frac{1}{p+1} - \frac{1}{p+\ell+1} \right]$$

and repeated integration by parts gives

$$\int_0^1 D^{K+1}[u(1-u)]^{K+1} du \int_u^1 D^{K+2}[v(1-v)]^{K+1} dv = - \frac{(2K+2)! \ (K+1)!^2}{(2K+3)!} .$$

Equating the two right hand sides, the coefficient of $R_{(K,K+1)}(\eta_i^{\cdot}, a_i^{\cdot})$ becomes

$$\frac{(K+1)!^2}{(2K+2)!(2K+3)!} \frac{1}{n^{2K+3}} .$$

Therefore

$$M_{i,i} = \frac{(K+1)!^2}{(2K+2)! \ (2K+3)!} \frac{1}{n^{2K+3}} \{ R_{(K+1,K)}(a_i^{\cdot}, \eta_i^{\cdot}) - R_{(K+1,K)}(a_i, \eta_i) \} .$$

Using $\frac{1}{n} = h(\zeta_i)(t_{i+1} - t_i)$, $t_i < \zeta_i < t_{i+1}$, we obtain by Riemann integrability,

$$n^{2K+2} \sum_i M_{i,i} \xrightarrow{n} \frac{(K+1)!^2}{(2K+2)! \ (2K+3)!} \int_a^b [R_{(K+1,K)}(t-,t) - R_{(K+1,K)}(t+,t)]h(t)dt$$

$$= \frac{(K+1)!^2}{(2K+2)! \ (2K+3)!} \int_a^b \frac{\phi^2(t) \ \alpha_K(t)}{h^{2K+2}(t)} dt$$

(by Lemma 2 in BC (1989)).

Off-diagonal terms $M_{i,j}$, $i \neq j$.

Integration by parts gives

$$\int_{t_i}^{t_{i+1}} P_{(1)}(t;t_i,t_{i+1}) \, R_{(K,K+1)}(t,s)h(t) \, dt$$

$$= P_{(0)}(t_{i+1};t_i,t_{i+1}) \, R_{(K,K+1)}(t_{i+1},s) - P_{(0)}(t_i;t_i,t_{i+1})R_{(K,K+1)}(t_i,s)$$

$$- \int_{t_i}^{t_{i+1}} P_{(0)}(t;t_i,t_{i+1})R_{(K+1,K+1)}(t,s)h(t) \, dt.$$

so the off-diagonal term becomes

$$M_{i,j} = - \frac{(K+1)!^2}{(2K+2)!^2} \left\{ \int_{t_j}^{t_{j+1}} [P_{(0)}(t_{i+1};t_i,t_{i+1})R_{(K,K+1)}(t_{i+1},s) \right.$$

$$- P_{(0)}(t_i;t_i,t_{i+1})R_{(K,K+1)}(t_i,s)] \, P_{(0)}(s;t_j,t_{j+1})h(s) \, ds$$

$$\left. - \int_{t_i}^{t_{i+1}}\int_{t_j}^{t_{j+1}} P_{(0)}(t;t_i,t_{i+1})P_{(0)}(s;t_j,t_{j+1})R_{(K+1,K+1)}(t,s)h(t)h(s) \, dt \, ds \right\}.$$

The first term has already been shown to be identically zero. Thus, we have

$$M_{i,j} = \frac{(K+1)!^2}{(2K+2)!^2} \int_{t_i}^{t_{i+1}} \int_{t_j}^{t_{j+1}} P_{(0)}(t;t_i,t_{i+1})P_{(0)}(s;t_j,t_{j+1})$$

$$\times R_{(K+1,K+1)}(t,s)h(t)h(s) \, dt \, ds.$$

$$= \frac{(K+1)!^2}{(2K+2)!^2} \sum_{\ell=0}^{K+1} \frac{(K+1+\ell)!(-1)^\ell}{(K+1-\ell)! \, \ell!^2} \frac{1}{n^{K+1-\ell}} \sum_{p=0}^{K+1} \frac{(K+1+p)!(-1)^p}{(K+1-p)! \, p!^2} \frac{1}{n^{K+1-p}}$$

$$\times \int_{t_i}^{t_{i+1}} \int_{t_j}^{t_{j+1}} H^\ell(t_i,t)H^p(t_j,s)R_{(K+1,K+1)}(t,s)h(t)h(s) \, dt \, ds.$$

Applying the Mean Value Theorem, we obtain

$$M_{i,j} = \frac{(K+1)!^2}{(2K+2)!^2} \frac{1}{n^{2K+4}} \sum_{\ell=0}^{K+1} \frac{(K+1+\ell)!(-1)^\ell}{(K+1-\ell)! \, \ell!(\ell+1)!}$$

$$\times \sum_{p=0}^{K+1} \frac{(K+1+p)!(-1)^p}{(K+1-p)! \, p!(p+1)!} R_{(K+1,K+1)}(a_{i,\ell},b_{j,p})$$

where $t_i < a_{i,\ell} < t_{i+1}$, $t_j < b_{j,p} < t_{j+1}$. It follows that

$$n^{2K+2} \sum_{i \neq j} M_{i,j} \xrightarrow[n]{} \frac{1}{(2K+2)!^2} \{ \int_0^1 D^{K+1}[u(1-u)]^{K+1} \, du \}^2$$

$$\times \int\int_{t \neq s} R_{(K+1,K+1)}(t,s)h(t)h(s) \, dtds = 0.$$

The final result is obtained by combining the diagonal and the off-diagonal terms. □

5. EXAMPLE

We consider the approximation of the integral

$$I = \int_0^1 e^{\beta t}X(t)dt,$$

where the stationary process X has covariance

$$R(t,s) = \{ 1 + \alpha|t-s| + \frac{\alpha^2}{3}(t-s)^2 \} e^{-\alpha|t-s|}$$

with spectral density $\varphi(\lambda) = (8\alpha^5/3\pi)(\alpha^2 + \lambda^2)^{-3}$, and thus exactly $K=2$ mean-square continuous q.m. derivatives. When $\beta \neq 0$, the density that generates the asymptotically optimal design is

$$h^*(t) = (2\beta/7)(e^{2\beta/7}-1)^{-1} e^{2\beta t/7}, \quad 0 \leq t \leq 1.$$

and the corresponding sampling points are

$$t^*_{i,n} = (7/2\beta) \ell n[1 + (e^{2\beta/7}-1)i/n], \quad i=0,\ldots,n.$$

The sample size of the design T_n is $N=n+1$. For simplicity of notation we will write t_i for $t_{i,n}$.

The simple-coefficient estimators (1.4), with $K=2$, which do not use the two existing q.m. derivatives, are denoted by I_n. The asymptotic constant in $n^6 E(I-I_n)^2 \rightarrow C_2$ has the following expression under the asymptotically optimal design

$$C_2^* = \frac{B_6}{6!} \{\int_0^1 [\alpha_2(t)e^{2\beta t}]^{1/7} dt \}^7 = \gamma^7 \alpha^5 / 5670 .$$

($\alpha_2(t) = 16\alpha^5/3$, $B_6 = 1/42$, $\gamma = (7/2\beta)(e^{2\beta/7}-1)$), and has the following expression under the uniform design

$$C_2^u = \frac{B_6}{6!} \int_0^1 \alpha_2(t)e^{2\beta t} dt = (e^{2\beta}-1)/11340.$$

The optimal-coefficient estimators (3.1) which use the existing two q.m. derivatives at the sample points $\{t_{i,n}^*\}_{i=0}^n$ are denoted by $_d\hat{I}_n^*$. The simple-coefficient estimators (4.1) which use the existing two q.m. derivatives at the sample points $\{t_{i,n}^*\}_{i=0}^n$ have the form:

$$_d I_n^* = \frac{1}{n} \{ \frac{1}{2} Y_{(0)}(0) + \sum_{i=1}^{n-1} Y_{(0)}(t_{i,n}^*) + \frac{1}{2} Y_{(0)}(1) \} + \frac{1}{10n^2} \{Y_{(1)}(0) - Y_{(1)}(1)\}$$

$$+ \frac{1}{60n^3} \{ \frac{1}{2} Y_{(2)}(0) + \sum_{i=1}^{n-1} Y_{(2)}(t_{i,n}^*) + \frac{1}{2} Y_{(2)}(1) \}$$

where

$$Y_{(0)}(t) = \gamma e^{(5\beta/7)t} X(t),$$

$$Y_{(1)}(t) = \gamma^2 e^{(3\beta/7)t} \{ \frac{5\beta}{49} X(t) + X'(t) \},$$

$$Y_{(2)}(t) = \gamma^3 e^{(\beta/7)t} \{ \frac{15\beta^2}{343} X(t) + \frac{26\beta}{49} X'(t) + X''(t) \}.$$

The asymptotic constant under the asymptotically optimal design in $n^6 E(I - _d\hat{I}_n^*)^2 \to _d C_2^*$ and in $n^6 E(I - _d I_n^*)^2 \to _d C_2^*$ is

$$_d C_2^* = \frac{1}{100800} \{\int_0^1 [\alpha_2(t)e^{2\beta t}]^{1/7} dt\}^7 = \gamma^7 \alpha^5 / 18900.$$

The simple-coefficient estimators (4.2) with K=2, which use the approximated existing two q.m. derivatives at the sample points $\{t_{i,n}^*\}_{i=0}^n$ are denoted by $_d\bar{I}_n^*$.

When β is close to zero, the asymptotically optimal sampling design

becomes close to periodic design. The improvement in performance of the asymptotically optimal design over the uniform design, when either estimators I_n^* or $_d\vec{I}_n^*$ are used, increases with β. We select a moderate value of $\beta = 3$ to differentiate between the two sampling designs.

For small values of α (highly correlated observations) the normalized mean square errors are very small; for example when $\alpha = 5$, a sampling design of size 3 gives a normalized m.s.e. of order 10^{-3}. For large values of α (weakly correlated observations) the normalized m.s.e.'s are significantly higher; for example when $\alpha = 20$, a sampling design gives a normalized m.s.e. of order .3.

The normalized m.s.e.'s corresponding to the optimal-coefficient estimators and simple-coefficient estimators that use the two existing q.m. derivatives, $E(I-_d\hat{I}_n^*)^2/EI^2$ and $E(I-_dI_n^*)^2/EI^2$, along with the asymptotic expression $n^{-6}{}_dC_2^*/EI^2$, are plotted versus the sample size $N=2,\dots,20$ in Figure 1 for $\alpha = 15$, $\beta = 3$. It is seen from Figure 1 that for small sample sizes the optimal-coefficient estimators $_d\hat{I}_n$ outperform (as expected) the simple-coefficient estimators $_dI_n^*$. However, for moderate n their performance is nearly identical and very close to their asymptotic performance.

The normalized m.s.e.'s corresponding to the two simple-coefficient estimators, $E(I-I_n)^2/EI^2$, $E(I-_d\bar{I}_n)^2/EI^2$, along with the asymptotic expression $n^{-6}C_2/EI^2$, are plotted versus the sample size $N=2,\dots,20$ in Figure 2 for $\alpha = 15$, $\beta = 3$, under both the asymptotically optimal (*) and the uniform (u) sampling designs. It is seen from Figure 2 that the asymptotically optimal design provides better performance than the uniform design. It is also seen that the m.s.e.'s of the two simple-coefficient estimators I_n and $_d\bar{I}_n$ are almost identical for all computed sample sizes, thus suggesting that $_d\bar{I}_n$ has the same asymptotic performance as I_n.

Comparing Figures 1 and 2, we see that the estimators that use the existing q.m. derivatives have better performance than those that do not for moderate and large sample sizes. Table 1 shows the number of samples N

required to achieve a specified performance under different designs and
estimators.

References

Benhenni, K. and Cambanis, S. (1989). Sampling designs for estimating integrals of stochastic processes. University of North Carolina Center for Stochastic Processes Technical Report No. 265. Apr. 89.

Eubank, R.L., Smith, P.L. and Smith, P.W. (1982). A note on optimal and asymptotically optimal designs for certain time series models. Ann. Statist. 10, 1295-1301.

Krylov, V.I. (1962). Approximate Calculation of Integrals. MacMillan, New York.

Sacks, J. and Ylvisaker, D. (1966). Designs for regression problems with correlated errors. Ann. Math. Statist. 37, 66-89.

Sacks, J. and Ylvisaker, D. (1968). Designs for regression problems with correlated errors; many parameters. Ann. Math. Statist. 39, 49-69.

Sacks, J. and Ylvisaker, D. (1970a). Designs for regression problems with correlated errors II. Ann. Math. Statist. 41, 2057-2074.

Sacks, J. and Ylvisaker, D. (1970b). Statistical designs and integral approximation. In: Proc. Twelfth Biennial Seminar of the Canadian Mathematical Congress, 115-136. Canadian Mathematical Congress, Montreal, Canada.

Shoenberg, I.J. (1969). Monosplines and quadrature formulae. In Theory and Applications of Spline Functions, T.N.E. Greville (ed.), Proc. Advanced Seminar (Math. Res. Center, Madison, Wis., 1968), MR 38 #3663.

Stein, M.L. (1988). Asymptotically efficient prediction of a random field with a misspecified covariance function. Ann. Statist. 16, 55-63.

Design	Estimator		Normalized m.s.e. 10^{-1}	10^{-2}	10^{-3}
Asymptotically optimal	Simple coefficients No derivatives	I_n^*	3	5	8
	Simple coefficients Approximated derivatives	$_d\overline{I}_n^*$	3	5	8
	From asymptotic expression No derivatives	(C_2^*)	4	5	8
Uniform	Simple coefficients No derivatives	I_n^u	4	7	11
	Simple coefficients Approximated derivatives	$_d\overline{I}_n^u$	4	7	11
	From asymptotic expression No derivatives	(C_2^u)	4	6	9
Asymptotically optimal	Optimal coefficients With derivatives	$_d\hat{I}_n^*$	2	4	6
	Simple coefficients With derivatives	$_dI_n^*$	3	4	6
	From asymptotic expression With derivatives	$(_dC_2^*)$	3	4	6

Table 1.   Number of samples N required to achieve a specified performance under different designs and estimators for $\alpha = 15$, $\beta = 3$.
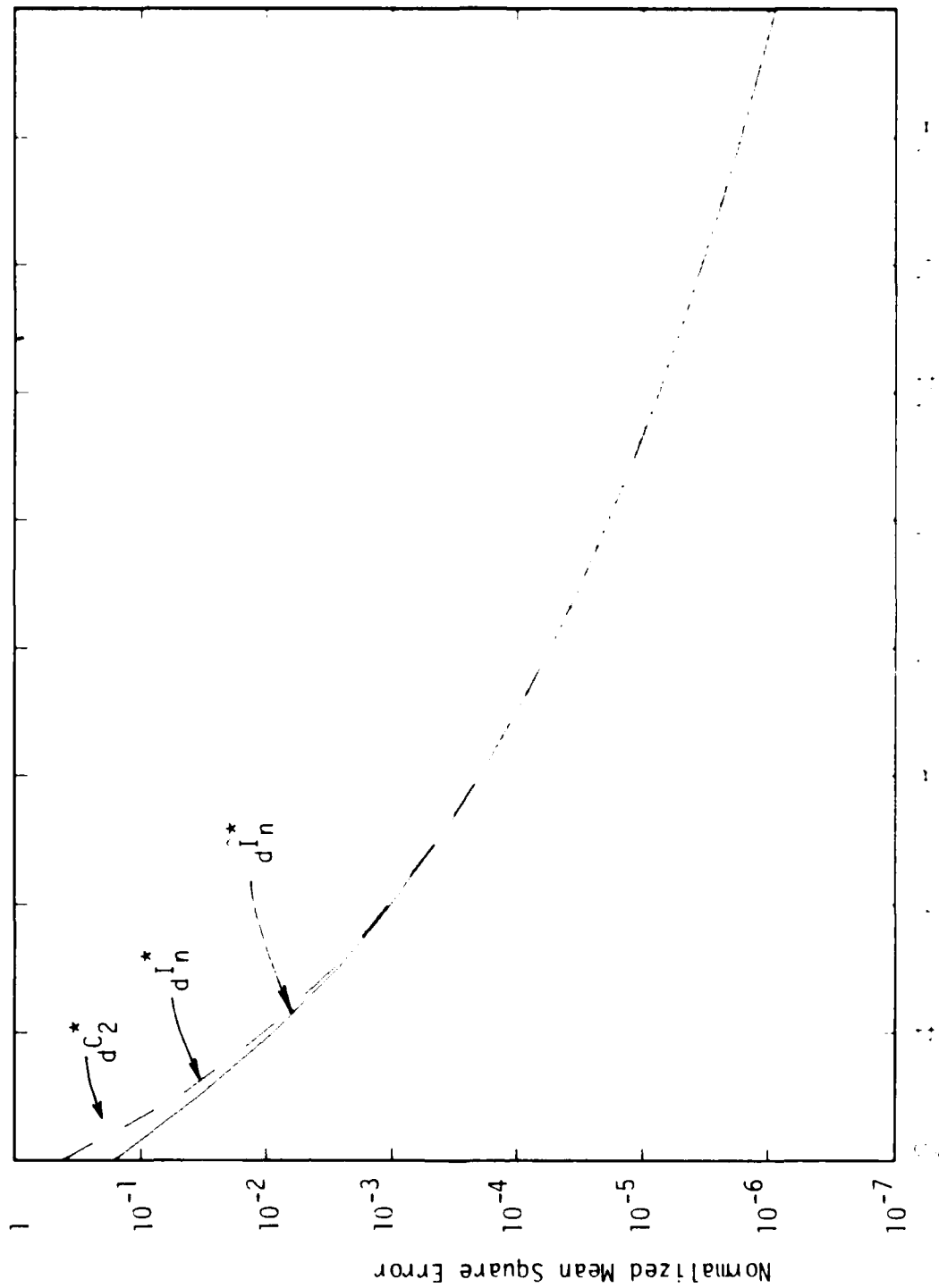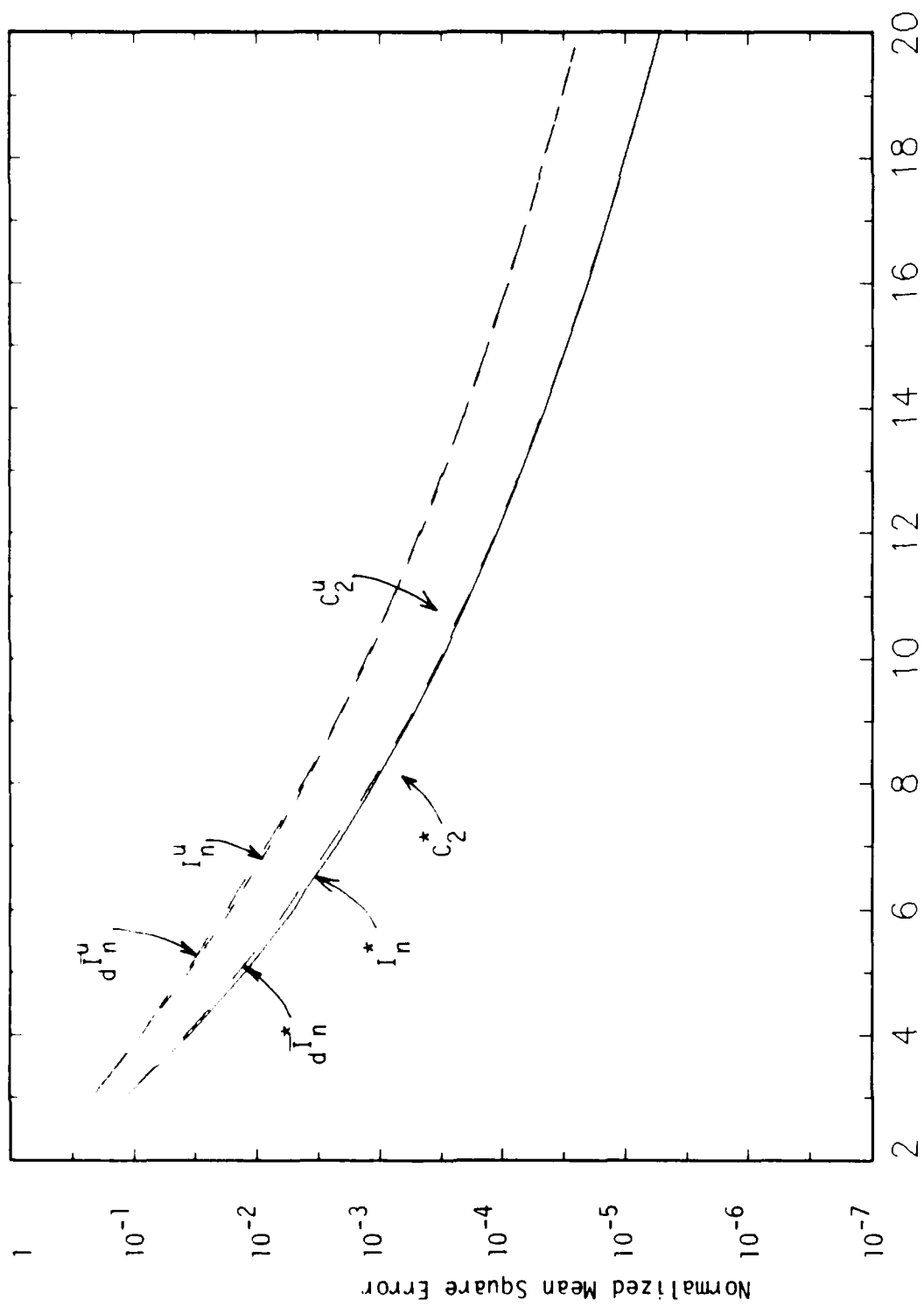
Sample Size (N)

Figure 1

Sample Size (N)

Figure 2

258. C. Houdré, Linear Fourier and stochastic analysis, Apr. 89.

259. G. Kallianpur, A line grid method in areal sampling and its connection with some early work of H. Robbins, Apr. 89. Amer. J. Math. Manag. Sci., 1989, to appear.

260. G. Kallianpur, A.G. Miamee and H. Niemi, On the prediction theory of two-parameter stationary random fields, Apr. 89. J. Multivariate Anal., 32, 1990, 120-149.

261. I. Herbst and L. Pitt, Diffusion equation techniques in stochastic monotonicity and positive correlations, Apr. 89.

262. R. Selukar, On estimation of Hilbert space valued parameters, Apr. 89. (Dissertation)

263. E. Mayer-Wolf, The noncontinuity of the inverse Radon transform with an application to probability laws, Apr. 89.

264. D. Monrad and W. Philipp, Approximation theorems for weakly dependent random vectors and Hilbert space valued martingales, Apr. 89.

265. K. Benhenni and S. Cambanis, Sampling designs for estimating integrals of stochastic processes, Apr. 89.

266. S. Evans, Association and random measures, May 89.

267. H.L. Hurd, Correlation theory of almost periodically correlated processes, June 89.

268. O. Kallenberg, Random time change and an integral representation for marked stopping times, June 89. Probab. Th. Rel. Fields, accepted.

269. O. Kallenberg, Some uses of point processes in multiple stochastic integration, Aug. 89.

270. W. Wu and S. Cambanis, Conditional variance of symmetric stable variables, Sept. 89.

271. J. Mijnheer, U-statistics and double stable integrals, Sept. 89.

272. O. Kallenberg, On an independence criterion for multiple Wiener integrals, Sept. 89.

273. G. Kallianpur, Infinite dimensional stochastic differential equations with applications, Sept. 89.

274. G.W. Johnson and G. Kallianpur, Homogeneous chaos, p-forms, scaling and the Feynman integral, Sept. 89.

275. T. Hida, A white noise theory of infinite dimensional calculus, Oct. 89.

276. K. Benhenni, Sample designs for estimating integrals of stochastic processes, Oct. 89. (Dissertation)

277. I. Rychlik, The two-barrier problem for continuously differentiable processes, Oct. 89.

278. G. Kallianpur and R. Selukar, Estimation of Hilbert space valued parameters by the method of sieves, Oct. 89.

279. G. Kallianpur and R. Selukar, Parameter estimation in linear filtering, Oct. 89.

280. P. Bloomfield and H.L. Hurd, Periodic correlation in stratospheric ozone time series, Oct. 89.

281. J.M. Anderson, J. Horowitz and L.D. Pitt, On the existence of local times: a geometric study, Jan. 90.

282. G. Lindgren and I. Rychlik, Slepian models and regression approximations in crossing and extreme value theory, Jan. 90.

283. H.L. Koul, M-estimators in linear models with long range dependent errors, Feb. 90.

284. H.L. Hurd, Almost periodically unitary stochastic processes, Feb. 90.

285. M.R. Leadbetter, On a basis for 'Peaks over Threshold' modeling, Mar. 90.

286. S. Cambanis and E. Masry, Trapezoidal stratified Monte Carlo integration, Mar. 90.

287. M. Marques and S. Cambanis, Dichotomies for certain product measures and stable processes, Mar. 90.

288. M. Maejima and Y. Morita, Trimmed sums of mixing triangular arrays with stationary rows, Mar. 90.

289. S. Cambanis and M. Maejima, Characterizations of one-sided linear fractional Lévy motions, Mar. 90.

290. N. Kôno and M. Maejima, Hölder continuity of sample paths of some self-similar stable processes, Mar. 90.

291. M. Merkle, Multi-Hilbertian spaces and their duals, Mar. 90

292. H. Rootzén, M.R. Leadbetter and L. de Haan, Tail and quantile estimation for strongly mixing stationary sequences, Apr. 90.

293. K. Benhenni and S. Cambanis, Sampling designs for estimating integrals of stochastic processes using quadratic mean derivatives, Apl. 90.

294. S. Nandagopalan, On estimating the extremal index for a class of stationary sequences, Apr. 90.

295. M.R. Leadbetter and H. Rootzén, On central limit theory for families of strongly mixing additive set functions, May 90.

296. W. Wu, E. Carlstein and S. Cambanis, Bootstrapping the sample mean for data from general distribution, May 90.

297. S. Cambanis and C. Houdré, Stable noise: moving averages vs Fourier transforms, May 90.

298. T.S. Chiang, G. Kallianpur and P. Sundar, Propagation of chaos and the McKean-Vlasov equation in duals of nuclear spaces, May 90.

299. J.M.P. Albin, On the upper and lower classes for stationary Gaussian fields on Abelian groups with a regularly varying entropy, June 90.